



PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN THE APPLICATION OF:

CONFIRMATION NO.: 3257

AKHILESWAR GANESH VAIDYANATHAN
ET. AL.

CASE NO.: CL1666USNA

APPLICATION NO.: 09/851674

GROUP ART UNIT: 1631

FILED: MAY 09, 2001

EXAMINER: LIN, JERRY

FOR: METHOD OF DISCOVERING PATTERNS IN SYMBOL SEQUENCES

APPEAL BRIEF UNDER 37 CFR 41.37

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the practice under 37 CFR 41.37, the following is a brief in support of the Appeal filed April 26, 2006, appealing the Rejection dated December 27, 2005 of Claims 35-53 and 66-68, all of the pending claims of this application.

A Petition for Extension of Time is enclosed with this paper.

Please charge any necessary Appeal Brief pursuant to 37 CFR 41.20(b)(2), to Deposit Account No. 04-1928 (E. I. du Pont de Nemours and Company). The Commissioner is hereby authorized to charge any additional fees which may be required or credit any overpayment to Deposit Account No. 04-1928.

05/09/2006 10:22:13S 00000002 041928 03851674
02 F0140E 500.00 00



TABLE OF CONTENTS

REAL PARTY IN INTEREST	3
RELATED APPEALS AND INTERFERENCES	4
STATUS OF CLAIMS.....	5
STATUS OF AMENDMENTS.....	6
SUMMARY OF CLAIMED SUBJECT MATTER.....	7
GROUND OF REJECTION TO BE REVIEWED ON APPEAL	10
ARGUMENT	
I. Rejection Under 35 U.S.C. § 101.....	11
II. Rejection Under 35 U.S.C. § 103.....	13
III. Conclusion.....	13
CLAIMS APPENDIX.....	14
EVIDENCE APPENDIX.....	19
RELATED PROCEEDINGS APPENDIX.....	20

REAL PARTY IN INTEREST

The real party in interest is E. I. du Pont de Nemours and Company (the “Assignee”), 1007 Market Street, Wilmington, Delaware 19898, to whom this application has been assigned, said assignment being recorded at Reel 012223, Frame 0540.

RELATED APPEALS AND INTERFERENCES

There are no other known appeals or interferences which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

STATUS OF CLAIMS

Claims 1-34 Cancelled

Claims 35-53 Rejected

Claims 54-65 Cancelled

Claims 66-68 Rejected

Claims 35-53 and 66-68 are set forth in the Claims Appendix.

STATUS OF AMENDMENTS

All amendments filed subsequent to the Final Rejection of April 15, 2005 have been entered.

A request for continuing examination was granted and the claims again rejected by the examiner by the paper dated December 27, 2005.

No amendments have been submitted following the latest rejection.

SUMMARY OF CLAIMED SUBJECT MATTER

In its method aspect recited in independent claims 35, 42 and 44 the present invention is directed to the identification of existing patterns in a set of "k" number of sequences. The k number of sequences is termed a "k-tuple". The k number of sequences form part of an overall set of "w" number of sequences. Each of the w sequences has a given length, but the sequences need not be the same length. A "pattern" is a distributed substring of elements that occurs in at least two sequences in a set of sequences (Page 7, lines 36-38)

The basic steps that comprise the core of the method of the present invention may be understood from the following discussion of a "two-tuple" ($k = 2$) of sequences S_1 and S_2 .

Each member element of a sequence is represented by an alphabetic symbol. Page 7, lines 30-31 and 33-34 show the two representative sequences S_1 and S_2 . Each symbol occupies a given location in a sequence. This location is termed the symbol's "position index". The pairing of a symbol and its position index identifies a unique symbol at a unique location in a sequence.

The first step of the method is to create for each sequence a table of ordered (symbol, position index) pairs. For instance, in the sequence S_1 the symbol "L" occurs at position indices 18 and 46 while the symbol "K" occurs at position indices 20, 25, 34 and 35. In the sequence S_2 the symbol "L" occurs at position indices 6, 23 and 30 and the symbol "K" occurs at position indices 8, 10, 14 and 32.

The association of each symbol and its position index is used to form a "master offset table" for each sequence. Figure 1 shows two master offset tables for the "two-tuple" of sequences S_1 and S_2 . Each master offset table groups, for each symbol, the position in the sequence occupied by each occurrence of that symbol.

Thus, in the master offset table for the sequence S_1 the position indices "18" and "46" are listed under the symbol "L" while position indices "20", "25", "34" and "35" are listed under the symbol "K". Similarly, for the sequence S_2 position indices "6", "23" and "30" are listed for the symbol "L" and position indices "8", "10", "14" and "32" are listed for the symbol "K".

Next, the difference-in-position between each occurrence of a symbol in one of the sequences and each occurrence of that same symbol in the other sequence is determined. This determination is facilitated by concatenating the two sequences. Page 10, line 15. A table, termed a "pattern map" (page 9, lines 32-33) or a "tuple-table" (page 30, line 25 through page 31, line 15), is formed in which each row in the table represents a single value of "difference in position" (page 9, lines 20 through page 10, line 6).

Figures 2A and 2B depict the pattern map for the two-tuple of sequences S_1 and S_2 . Since sequence S_1 contains 47 characters and the sequence S_2 contains 54 characters the pattern map is 100 rows in depth. For each given value of a difference-in-position (the value being termed the "row index") the table lists the position of each symbol in the first sequence that appears again at a spacing corresponding to that difference-in-position value.

Consider the symbol "R" listed in the master offset table for the sequence S_1 (at position index "44") and the position indices for the same symbol "R" as listed in the master offset table for the sequence S_2 (position indices "7", "21", "31"). From the master offset tables and the concatenation of the sequences S_1 and S_2 at page 7 it may be determined that:

- from the occurrence of the symbol "R" in the first sequence,
 - the first occurrence of the symbol "R" in the second sequence is spaced ten places;
 - the second occurrence of the symbol "R" in the second sequence is spaced twenty-four places; and
 - the third occurrence of the symbol "R" in the second sequence is spaced thirty-four places.

The pattern map of Figures 2A and 2B thus lists the position index "44" (corresponding to the symbol "R") on row indices (difference-in-position values) "10", "24" and "34".

The symbols collected for any row index (each value of difference-in-position) define a parent pattern in the first sequence that is repeated in the second sequence.

Consider the discussion at page 11, line 24 through page 12, line 30 for the row index value "35" in the pattern map of Figure 2A. This row index value identifies the pattern corresponding to the symbols at position indices "18", "20", "21", "30", "39" and "40". [The value "6" in the symbol count column to the immediate right of the colon on Figure 2A (page 10, line 37 through page 11, line 2) indicates that there are six symbols in the pattern.]

By consulting sequence S_1 the position indices "18", "20", "21", "30", "39" and "40" respectively correspond to the symbols "L", "K", "V", "V", "P", "H".

The collected symbols corresponding to a difference-in-position value "35" thus identifies the pattern occurring in the first sequence S_1 as:

"L . KV V PH"

that also appears in the second sequence S_2 (page 12, line 16), where the dots indicate placeholders in the pattern (page 12, lines 19-21).

Claim 66 is directed a computer-readable medium containing a data structure useful by a computer system in the practice of the method steps described above.

Claim 68 is directed to a computer-readable medium containing instructions for controlling a computer system to discover one or more patterns in two sequences of symbols S_1 and S_2 by performing the method steps described above.

Both claims 66 and 68 contain language using difference-in-position values as the selection criteria for identifying repeating patterns of symbols.

GROUND OF REJECTION TO BE REVIEWED ON APPEAL

Claims 35-53 and 66-68 are rejected under 35 U.S.C. § 101 as being directed to non-statutory subject matter.

Claims 66 and 67 are rejected under 35 U.S.C. § 103(a) as being unpatentable over United States Patent 5,577,249 (Califano).

ARGUMENT

I. Rejection Under 35 U.S.C. § 101

Claims 35-53 and 66-68 are rejected as being presented to non-statutory subject matter. Since the present invention as recited in those claims operates to identify and to select patterns of alphabetic symbols present in a set of one or more symbol sequences these claims define an invention that is sufficiently “useful, tangible and concrete” to qualify as statutory subject matter under 35 U.S.C. § 101.

As recited in the claims, the invention discovers patterns in a set of sequences of symbols. The claims clearly recite that symbols are members of an alphabet.

On one level, an alphabetic character whether represented in print (as a collection of ink molecules on a substrate) or electronically (as a collection of electrons) is, in and of itself, a physical thing.

A sequence, as a string of such characters, is also, in and of itself, a physical thing. A “sequence” cannot exist in isolation. A “sequence” must have “things” in it.

On still another level, an alphabetic symbol (itself a “thing”) represents other “things” that are themselves physical objects, whether those “other things” be chemical components of a protein or an amino acid, an amount of dollars or shares, or any other physical entity.

Clearly, on any of these levels the claimed method can be construed to operate on tangible and concrete objects -- either the characters themselves, the sequences of the characters, or the physical objects represented by the characters.

Moreover, the operation is not an abstract organizational effort. At the end of the day, if a pattern is present in two or more sequences, the practice of the present invention results in the identification of that pattern. The identification of patterns of symbols that occur within plural sequences and which contain information about physical things is sufficiently “useful, tangible and concrete” to qualify as statutory subject matter under 35 U.S.C. § 101. State Street Bank & Trust Co. v. Signature Financial Group, Inc., 47 USPQ2d 1596 (Fed. Cir. 1998).

The examiner is incorrect when he asserts that present invention is a “mathematical algorithm that analyses symbols”.

It is clearly the law that even though a mathematical algorithm may not itself be eligible subject matter, the practical use of that algorithm to produce a useful, tangible and concrete result is. In re Alappat, 31 USPQ2d at 1578-79 (Fed. Cir. 1994).

The examiner's characterization of the invention is erroneous. The invention does not "organiz[e] patterns among abstract symbols" or "analyse[s] symbols". The invention does not seek "merely [to] organize, sort, and arrange data".

Any pattern is already inherently present in the sequence itself. The symbols are not manipulated to create any patterns. The symbols that comprise the patterns are not themselves "analyzed".

What the present invention does do is provide a method whereby any patterns that inherently exist in sequences are identified. This identification is a useful, tangible and concrete result with practical value.

The examiner also contends that the invention does not "select any data" (Office Action dated April 15, 2005, page 4; reiterated and maintained in the rejection of December 27, 2005, pages 2 and 5.)

The examiner apparently ascribes some importance to the term "select". [It is noted in passing that it is perhaps the express use of that term that renders statutory the subject matter claimed in the Califano reference (U.S. Patent 5,577,249) relied upon by the Examiner, since both Califano and the present invention find or discover some thing from among a group of things. Yet Califano is apparently directed to statutory subject matter while the examiner contends that the present invention is not.]

In any event, this contention is also erroneous, because the present invention does contain the equivalent of a "selection" step (rejection of December 27, 2005, page 4.)

Webster's Third New International Dictionary of the English Language Unabridged, Merriam-Webster Inc., Springfield, Massachusetts 1961, at page 444 (a copy of which is appended to this paper for the convenience of the Board) defines "collect" to mean:

"4 : to bring together esp. in accordance with a principle of selection or an informative or profitable end" (emphasis added).

Clearly, defining patterns that exist in the sequences by "collecting" symbols in accordance with identical difference-in-position values is a "selection".

Each of the independent claims 35, 42, 44, and 68 defines patterns by "collecting" symbols in accordance with a predetermined criteria (identical difference-in-position value). Claim 66 uses that same criteria to the same effect. Thus, all of these claims clearly contain a "selection" step.

In sum, the identification of patterns of symbols occurring within plural sequences is sufficiently “useful, tangible and concrete” to qualify as statutory subject matter under 35 U.S.C. § 101.

II. Rejection under 35 U.S.C. § 103

Claims 66-67 have been rejected as unpatentable under 35 U.S.C. § 103(a) as evidenced by U.S. Patent 5,577,249 (Califano).

Claim 66 includes language that clearly distinguishes the claimed data structures from Califano.

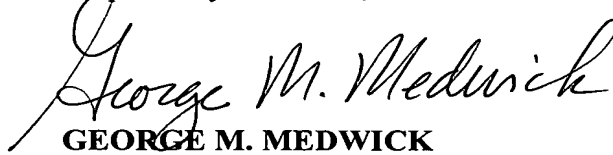
The data structure of Califano does not include and does not utilize identical difference-in-position values to define patterns of symbols.

Since a data structure in accordance with the present invention includes this relationship between difference-in-position values and patterns, the present invention is not rendered obvious by the Califano reference.

III. Conclusion

For the reasons set forth it is submitted that the Examiner's rejections are improper and should be reversed, which action is earnestly solicited.

Respectfully submitted,



GEORGE M. MEDWICK
ATTORNEY FOR ASSIGNEE
Registration No.: 27,456
Telephone: (302) 892-7915
Facsimile: (302) 992-5374

Dated: Sept. 25, 2006

CLAIMS APPENDIX

Claims 1-34 Cancelled

35. (Previously presented) A method of discovering one or more patterns in a set of k sequences of symbols, called a k -tuple, where k is greater than or equal to two, within an overall set of w sequences having sequence numbers 1, 2, ..., w , the symbols being members of an alphabet, each sequence of symbols having respective lengths L_1, L_2, \dots, L_w , comprising the steps of:

a) translating the sequences of symbols into a table of ordered (symbol, position index) pairs, where the position index refers to the location of the symbol in a sequence;

b) for each of the w sequences, grouping the (symbol, position index) pairs by symbol to form a respective master offset table, thus creating w master offset tables;

c) using the position indices in the w master offset tables to determine the difference-in-position value between each occurrence of a symbol in one of the sequences and each occurrence of that same symbol in the other sequence in each master offset table, forming a k -tuple table associated with the k -tuple, the table comprising k columns, one of the k columns being a primary column and the remaining $(k-1)$ columns being suffix columns, each column corresponding to one of the k sequences;

i) the primary column comprising the (symbol, position index) pairs of a primary sequence,

ii) the $(k-1)$ suffix columns comprising (symbol, difference-in-position value) pairs, where the difference-in-position values are the position differences between all same symbols of each remaining sequence of the tuple and the primary sequence of the tuple,

iii) the rows in the k -tuple table resulting from forming all combinations of same symbols from each sequence;

d) creating a sorted k -tuple table by performing a multi-key sort on the k -tuple table, the sort keys being selected respectively from the difference-in-position values of the last suffix column (k^{th} column) through the difference-in-position value of the first suffix column; and

e) defining one or more patterns by collecting adjacent rows of the sorted k -tuple table whose suffix columns contain identical difference-in-position values, the relative positions of the symbols in each pattern being determined by the primary column position indices, the one or more patterns being common to the k sequences.

36. (Original) The method of claim 35 further comprising:

f) deleting all patterns not satisfying a predetermined criteria.

37. (Original) The method of claim 35 further comprising:

f) deleting all patterns shorter than a first predetermined span and longer than a second predetermined span.

38. (Original) The method of claim 35 further comprising:

f) deleting all patterns having fewer than a predetermined number of symbols.

39. (Original) The method of claim 35, further comprising the step of deleting rows from the k -tuple table which do not have suffix indices identical to any other row of the k -tuple table.

40. (Original) The method of claim 35 further comprising the step of deleting rows from the k-tuple table according to predetermined criteria.

41. (Previously presented) The method of claim 40, wherein rows sharing identical suffix column difference-in-position values are deleted from the k-tuple table if there are fewer than N_s such rows, where N_s is the minimum number of symbols per pattern.

42. (Previously presented) A method of discovering one or more patterns in a set of $k+1$ sequences of symbols, called a $(k+1)$ -tuple, where k is greater than or equal to two, within an overall set of w sequences having sequence numbers 1, 2, ..., w , the symbols being members of an alphabet, each sequence of symbols having respective lengths L_1, L_2, \dots, L_w by first forming a k-tuple table and then forming a $(k+1)$ -tuple table by combining the k-tuple table with an additional sequence of symbols, the formation of the k-tuple table comprising the steps of:

a) translating the sequences of symbols into a table of ordered (symbol, position index) pairs, where the position index refers to the location of the symbol in a sequence;

b) for each of the w sequences, grouping the (symbol, position index) pairs by symbol to form a respective master offset table, thus creating w master offset tables;

c) using the position indices in the w master offset tables to determine the difference-in-position value between each occurrence of a symbol in one of the sequences and each occurrence of that same symbol in the other sequence in each master offset table, forming a k-tuple table associated with the k-tuple, the table comprising k columns, one of the k columns being a primary column and the remaining $(k-1)$ columns being suffix columns, each column corresponding to one of the k sequences;

i) the primary column comprising the (symbol, position index) pairs of a primary sequence,

ii) the $(k-1)$ suffix columns comprising (symbol, difference-in-position value) pairs, where the difference-in-position values are the position differences between all same symbols of each remaining sequence of the tuple and the primary sequence of the tuple,

iii) the rows in the k-tuple table resulting from forming all combinations of same symbols from each sequence;

d) creating a sorted k-tuple table by performing a multi-key sort on the k-tuple table, the sort keys being selected respectively from the difference-in-position values of the last suffix column (k^{th} column) through the difference-in-position value of the first suffix column; the formation of the $(k+1)$ -tuple table comprising the steps of:

e) translating the additional sequence of symbols into a table of ordered (symbol, position index) pairs, where the position index refers to the location of the symbol in the additional sequence of symbols;

f) grouping the (symbol, position index) pairs by symbol to form a master offset table;

g) creating the $(k+1)$ -tuple table of $k+1$ columns, one of the $k+1$ columns being a primary column and the remaining k columns being suffix columns, by:

i) forming all combinations of same symbols between the primary column of the k-tuple table and the master offset table,

ii) for each such combination, duplicating the corresponding row of the k-tuple table, and appending a (symbol, difference-in-position value) pair corresponding to the difference between the position index of the master offset table and the position index of the primary column;

h) creating a sorted $(k+1)$ -tuple table by performing a multi-key sort on the $(k+1)$ -tuple table, the sort keys being selected respectively from the difference-in-position values of the

last suffix column $[(k+1)^{\text{th}}$ column] through the difference-in-position value of the first suffix column; and

i) defining one or more patterns by collecting adjacent rows of the sorted $(k+1)$ -tuple table whose suffix columns contain identical difference-in-position values, the relative positions of the symbols in each pattern being determined by the primary column position indices, the one or more patterns being common to the $k+1$ sequences.

43. (Previously presented) The method of claim 42 further comprising the step of: deleting patterns from a k -tuple table common to the k -tuple table and a $(k+1)$ -tuple table, where the $(k+1)$ -tuple table contains all of the sequences of the k -tuple table with one additional sequence, by:

a) deleting the suffix column corresponding to a sequence not shared between the two tuple tables, thereby defining a modified table, and

b) deleting all rows from the k -tuple table whose suffix columns all contain identical sets of difference-in-position values to a row of the modified table.

44. (Previously presented) A method of discovering one or more patterns in a set of k sequences of symbols, called a k -tuple, comprising the steps of:

a) for a first pair of sequences of the k -tuple

i) translating each sequence of symbols into a table of ordered (symbol, position index) pairs, where the position index of each (symbol, position index) pair refers to the location of the symbol in the sequence;

ii) for each of the paired sequences, grouping the (symbol, position index) pairs by symbol to respectively form a first master offset table and a second master offset table;

iii) forming a Pattern Map comprising an array having $(L1 + L2 - 1)$ rows by:

A) subtracting the position index of the first master offset table from the position index of the second master offset table for every combination of (symbol, position index) pair having same symbols, the difference-in-position value resulting from each subtraction defining a row index;

B) storing each (symbol, position index) pair from the first master offset table in a row of the Pattern Map, the row being defined by the row index, until all (symbol, position index) pairs have been stored in the Pattern Map;

iv) defining a parent pattern by collecting symbols having the identical difference-in-position value from each row of the Pattern Map and populating an output array with the collected symbols, the symbols being placed at relative locations in the parent pattern indicated by the position index of the (symbol, position index) pair; and

v) repeating step iv) for each row of the Pattern Map;

b) storing the discovered patterns as arrays of (symbol, position index) pairs;

c) for each subsequent pair of sequences of the k -tuple, replacing the (symbol, position index) pairs of the first sequence of the pair of sequences by the (symbol, position index) pairs of the stored patterns; and

d) repeating steps (a) through (c) for each subsequent pair of sequences until the k -th sequence of the k -tuple is reached.

45. (Previously presented) The method of claim 35, further comprising the step of finding all patterns at all levels of support within a set of sequences by:

f) forming a tree of nodes, where each node corresponds to each combination of k sequences, and therefore represents a k-tuple, and wherein each node representing a k-tuple is connected to all nodes representing (k+1)-tuples,

each (k+1)-tuple being formed by adding a unique sequence to the k-tuple, where the sequence being added is later in the ordered list of sequences than the latest sequence of the k-tuple;

g) traversing the tree, and at each node visited during traversal, defining one or more patterns by collecting adjacent rows of the sorted k-tuple table whose suffix columns contain identical sets of difference-in-position values, the relative positions of the symbols in each pattern being determined by the primary column position indices, the one or more patterns being common to the k sequences.

46. (Original) The method of claim 45, wherein the traversal of the tree of nodes is accomplished via recursion.

47. (Previously presented) The method of claim 45, further comprising the step of:
h) removing duplicate patterns at each level of support.

48. (Previously presented) The method of claim 47, wherein the removal of duplicate patterns at each level of support step h) is accomplished by:

i) for each node corresponding to a (k+1)-tuple, identifying the nodes containing k-tuples whose sequences are subsets of the (k+1)-tuple; thereby defining a set of causally-dependent nodes;

ii) locating said causally-dependent nodes;

iii) removing from each said causally-dependent node the patterns in common with the (k+1)-tuple; and

iv) if the k-tuple table in a causally-dependent node is thereby reduced to zero length, removing the corresponding node and all of its descendents from the tree prior to their traversal.

49. (Previously presented) The method of claim 48, wherein locating causally-dependent nodes in step ii) comprises the steps of:

(A) organizing the nodes at level k in the Tuple-tree into a linked list which is ordered from left to right in accordance with the sequence numbers represented by each tuple; and

(B) searching said linked list for nodes which are causally-dependent on a particular (k+1)-tuple.

50. (Original) The method of claim 48, wherein the nodes located in step ii) are causally-dependent nodes at level k determined with respect to another node at level k, and are thus causally-dependent on a child of the another node at level k.

51. (Previously presented) The method of claim 47, wherein the removal of duplicate patterns at each level of support step h) comprises the steps of:

i) organizing the nodes at level k in the Tuple-tree into a linked list which is ordered from left to right in accordance with the sequence numbers of each tuple;

ii) for each pattern in the current node at level k, storing a "hit list" array of the sequence numbers of the sequences containing the pattern;

iii) for all nodes to the right of the current node whose sequence numbers are all in the hit list, searching for a duplicate instance of the pattern, and if found, eliminating it; and

iv) making each node the current node, repeating steps (ii) and (iii), in the order of the node's appearance in the linked list.

52. (Previously presented) The method of claim 51, wherein, in step iii), the nodes consistent with the hit list are found using a hash tree, the hash tree having a root and k levels of nodes, the k-th level of the hash tree having a plurality of leaf nodes, the respective level of nodes of the hash tree corresponding to the respective sequence numbers of a k-tuple, the leaf nodes identifying the k-tuple whose sequence numbers correspond to the path from the root to the leaf node, wherein

searching the nodes for pattern duplicates is performed by repeating steps ii) and iii) for each node in the order of the appearance of that node in the hash tree.

53. (Previously presented) The method of claim 45 wherein the traversing step c) itself comprises the steps of:

i) creating a Virtual Sequence Array of patterns found within the sequences, wherein the patterns are termed P-nodes and the tuple nodes are termed T-nodes,

(ii) finding a P-node list corresponding to the location of each pattern in the primary sequence of that tree node,

iii) searching the P-node list for a duplicate instance of the pattern,

(A) if no duplicate is found:

(1) adding a pointer to the pattern of the current T-node pattern array,

(2) finding all locations of the pattern within the Virtual Sequence Array,

(3) adding a pointer to the pattern to each corresponding P-node array;

(B) if a duplicate pattern is found:

(1) ignoring the pattern if the duplicate pattern was found at support equal to the current level of support,

(2) if the duplicate pattern was found at a previous level of support, unlinking the duplicate pattern from its previous T-node (if it exists), and relinking the duplicate pattern to the current T-node,

(3) repeating steps 1) and 2) until all of the children of a T-node have been created, thus insuring that patterns of that T-node that are at their ultimate level of support are reported, and

(4) deleting the T-node.

Claims 54-65 (Cancelled)

66. (Previously presented) A computer-readable medium containing a plurality of data structures useful in controlling a computer system to discover one or more patterns in k sequences of symbols within an overall set of w sequences, the plurality of data structures comprising:

a number w of master offset table data structures each grouping,

for each value of a difference in position between each occurrence of a symbol in one of the sequences and each occurrence of that same symbol in each other sequence,

the position (position index) in the first sequence of each symbol therein that appears in each of the other sequences at that difference-in-position value;

a k-tuple table data structure comprising columns and rows, the columns comprising (symbol, position index) pairs and (symbol, difference-in-position value) pairs; and

a sorted k-tuple table data structure comprising a row-sorted representation of the (symbol, position index) pairs and (symbol, difference-in-position value) pairs contained in the k-tuple table data structure,

wherein adjacent rows of the sorted k-tuple table data structure whose suffix columns contain identical difference-in-position values define one or more patterns of symbols, the relative positions of symbols in each pattern being determined by the primary column position indices in the sorted k-tuple table data structure.

67. (Previously presented) The computer-readable medium of claim 66 wherein the sorted k-tuple table data structure further groups, for each value of a difference in position, the number of symbols in the first sequence that appear in the second sequence at that difference-in-position value.

68. (Previously presented) A computer-readable medium containing instructions for controlling a computer system to discover one or more patterns in a set of k sequences of symbols, called a k-tuple, where k is greater than or equal to two, within an overall set of w sequences having sequence numbers 1, 2, ..., w, the symbols being members of an alphabet, each sequence of symbols having respective lengths L_1, L_2, \dots, L_w , by executing a method comprising the steps of:

a) translating the sequences of symbols into a table of ordered (symbol, position index) pairs, where the position index refers to the location of the symbol in a sequence;

b) for each of the w sequences, grouping the (symbol, position index) pairs by symbol to form a respective master offset table, thus creating w master offset tables;

c) using the position indices in the w master offset tables to determine the difference-in-position value between each occurrence of a symbol in one of the sequences and each occurrence of that same symbol in the other sequence in each master offset table, forming a k-tuple table associated with the k-tuple, the table comprising k columns, one of the k columns being a primary column and the remaining (k-1) columns being suffix columns, each column corresponding to one of the k sequences;

i) the primary column comprising the (symbol, position index) pairs of a primary sequence,

ii) the (k-1) suffix columns comprising (symbol, difference-in-position value) pairs, where the difference-in-position values are the position differences between all same symbols of each remaining sequence of the tuple and the primary sequence of the tuple,

iii) the rows in the k-tuple table resulting from forming all combinations of same symbols from each sequence;

d) creating a sorted k-tuple table by performing a multi-key sort on the k-tuple table, the sort keys being selected respectively from the difference-in-position values of the last suffix column (k^{th} column) through the difference-in-position value of the first suffix column; and

e) defining a one or more patterns by collecting adjacent rows of the sorted k-tuple table whose suffix columns contain identical difference-in-position values, the relative positions of the symbols in each pattern being determined by the primary column position indices, the one or more patterns being common to the k sequences.

EVIDENCE APPENDIX

None

RELATED PROCEEDINGS APPENDIX

None

Webster's Third New International Dictionary

OF THE ENGLISH LANGUAGE
UNABRIDGED

A Merriam-Webster

REG. U.S. PAT. OFF.

*Utilizing all the experience and resources of more than
one hundred years of Merriam-Webster dictionaries*

EDITOR IN CHIEF

PHILIP BABCOCK GOVE, Ph.D.

AND

THE MERRIAM-WEBSTER

EDITORIAL STAFF



G. & C. MERRIAM COMPANY, *Publishers*

SPRINGFIELD, MASSACHUSETTS, U.S.A.

BEST AVAILABLE COPY

COPYRIGHT © 1961 BY G. & C. MERRIAM CO.

PREVIOUS EDITIONS

COPYRIGHT © 1909, 1913, 1923, 1924, 1926, 1927, 1930, 1934, 1953, 1957, 1959
BY G. & C. MERRIAM CO.

PHILIPPINES COPYRIGHT 1961 BY G. & C. MERRIAM CO.

PREVIOUS EDITIONS PHILIPPINES COPYRIGHT 1950, 1953, 1954, 1957, 1959

BY G. & C. MERRIAM CO.

PREVIOUS EDITION COPYRIGHT 1925 IN PHILIPPINE ISLANDS

BY G. & C. MERRIAM CO.

BEST AVAILABLE COPY

ALL RIGHTS RESERVED

UNDER INTERNATIONAL AND PAN-AMERICAN COPYRIGHT CONVENTIONS
BY G. & C. MERRIAM CO.

All rights reserved

MADE IN THE U. S. A.

R. R. DONNELLEY & SONS COMPANY, THE LAKESIDE PRESS, CHICAGO, ILL., U. S. A.

COMPOSITORS

H. O. HOUGHTON AND COMPANY, THE RIVERSIDE PRESS, CAMBRIDGE, MASS., U. S. A.

PRINTERS AND BINDERS

BEST AVAILABLE COPY